**Excel – Speed and Weather Analysis**

This activity is done in Excel first, and then lateryou can attempt it again in quicksight so that you already know what results to expect, and concentrate on learning how to use quicksight features rather than thinking if you have completed exercise correctly.

1. Activity 2 task (about 30 min)
2. Hints to give within 5…10 minutes:
   1. Use VLOOKUP to join three tables (XLOOKUP in the future?)
   2. Use PIVOT TABLES to answer the questions
3. When starting discussing the results, there will be two remarks:
   1. Someone will discover that we need to define «the most dangerous speed». The most acceptable definition is: «Highest probability of being involved in a fatal or a serious accident»
   2. Someone will discover that dataset does not contain the actual speed at the moment of an accident. Does it make this dataset useless? No. We know that people do speed. But speeding is usually correlated to the speed limit. The dataset is still relevant and can be used to answer the question. We make an assumption that ON AVERAGE cars were travelling at the speed close to the limit (even if slightly above – but still correlated to it)
4. There will also be two points related to data cleansing:
   1. What is that accident at 10 mph? Probably, car park or a private road. Naturally, slight accident only
   2. What is accident at 0 mph? That does not mean that the car was stationary – that is nominal speed limit, not the actual speed car was doing. However, there is no such road – probably, missing data
   3. It is safe to delete these records because of two reasons:
      1. They represent only 3 records out of 140k, so very small percentage
      2. Accident severity distribution amongst these records is exactly the same as for the whole dataset: most accidents are slight.
   4. Because of two factors above, it is safe to delete these records from our dataset
5. Question 1
6. Wrong conclusions
   1. To improve safety, let’s increase speed limit from 60 to 70 –because 70 is safer than 60. And don’t argue – data confirms that! As we know, numbers don’t lie.
      1. Someone will mention that 70 is usually found on a motorway, where there is central reservation – that what makes it safer comparing to other roads. It eliminates danger of head-on collision.
      2. Safety crash tests are performed at 56 mph. Two cars travelling in 30-miles zone at 35 mph yield collision at 70 mph – which is way above speed of crash tests. That’s why head-on is so dangerous – and that’s what makes motorways safer roads.
      3. Some time ago people used to say «motorways are usually safer but if something happens it is usually fatal» – is this still the case? Probably, not.
   2. Why 20 is more dangerous than 30? Yet again, should we increase speed limit to 30?
      1. Answer: 20-miles roads used to be 30-miles roads, but fatalities back then used to be much higher than the current value. After speed limit was decreased to 20, it went down significantly – even though it may still be a bit higher than in 30 zone. Conclusion: be careful in 20-miles zones, it is 20 for a reason!
7. Model answer workout:
   1. Open Accidents\_2015 dataset in Excel
   2. Open two more datasets (accident\_severity and weather\_conditions) and copy their content onto another sheet
   3. Link those values using VLOOKUP
   4. Optionally, can now copy the whole sheet and paste as values (to remove formulas)
8. Build pivot table. It is worth investing your time into learning about pivot tables – it is a very useful tool for analysts!
   1. Click anywhere within the table, excel will pick up on the table
   2. Insert into new worksheet
   3. Drag Accident severity into VALUES (it will change to count – that is counting unique facts, so must use primary key for that)
   4. Drag speed limit into ROWS
   5. Drag accident severity label into COLUMNS
   6. Check that the total in the pivot matches the number of records in the dataset - testing
9. Build the graph from it
   1. Insert recommended chart – it shows useless data
   2. We want to find a proportion between fatal, serious and slight accidents. So need to switch data in pivot table
   3. Right-click any value, and then click Show values as – Percent of Row total
   4. Change the graph type to stacked bar chart – it will add up fatal and serious accidents together
   5. From here, you can see the answer: 60
10. Answer: 60
11. Wrong graphs
    1. Open WRONG\_Speed tab
    2. Values are shown as absolute figures, and we can conclude that the most dangerous speed is 30. But this is only because majority of roads have this limit, and therefore majority of accidents happen there. Short trips entirely happen within 30-miles roads and never even reach motorways.
    3. Also, it would be useful to have the total number of people driving on the roads – but we do not have this information
    4. You can also compare data for 30 vs 60: total number of accidents is 4.5 times higher (90k vs 20k) while amount of fatal accident is higher for 60 (576 vs 523). This is not shown at all on the wrong graph!
12. Conclusion: that is why knowing the nature of the data, and understanding the meaning behind the data is so important
13. Save it as Excel workbook so the other sheets are not lost
14. Question 2
15. Working out is similar: build the same pivot table (drag weather\_conditions\_label instead of speed\_limit) and display values as percent of row total. Build the graph
16. Answers: No (it does make difference however, it is not significant)
17. What is the safest driving conditions? Surprisingly, snow – people drive carefully
18. Question 3 answer: Fog and mist (from the same graph)
    1. How to avoid: keep your fashion distance (main reason for car accidents in the UK)
19. Wrong conclusions (open WRONG\_WEATHER tab)
    1. The most dangerous weather conditions is fine with no winds – simply because majority of accidents happen at those conditions
    2. The real most dangerous conditions (fog and mist) are not even displayed (so small comparing to amount of accidents under fine weather)
20. To summarise, what we have done in Excel:
    1. VLOOKUP – equivalent of relationships in databases
    2. Pivot tables (there is a very good MS course on Excel)